VOLUME VII: MONTE-CARLO COMPARISON OF CONFIDENCE
INTERVAL PROCEDURES FOR ESTIMATING SURVIVAL IN A
RELEASE-RECAPTURE STUDY, WITH APPLICATIONS
TO SNAKE RIVER SALMONIDS

Technical Report

This document should be cited as follows:

This report and other BPA Fish and Wildlife Publications are available on the Internet at:

http://www.efw.bpa.gov/cgi-bin/efw/FW/publications.cgi

For other information on electronic documents or other printed media, contact or write to:

Bonneville Power Administration
Environment, Fish and Wildlife Division
P.O. Box 3621
905 N.E. 11th Avenue
Portland, OR 97208-3621

Please include title, author, and DOE/BP number in the request.

THE DESIGN AND ANALYSIS OF SALMONID TAGGING STUDIES IN THE COLUMBIA BASIN

VOLUME VII

Monte-Carlo Comparison of Confidence Interval Procedures for Estimating

Survival in a Release-Recapture Study, with Applications to Snake River Salmonids

Prepared by:

Alan B. Lowther
John R. Skalski

School of Fisheries
University of Washington
Seattle, Washington

PREFACE

Project 89-107, Epidemiological Survival Methods, was developed to provide statistical guidance on the design and analysis of PIT-tag survival studies to the Northwest fisheries community. Studies under this project have determined the statistical feasibility of conducting PIT-tag smolt survival studies, assessed analytical capabilities for analyzing the tagging experiments, and made recommendations on study design. As PIT-tag capabilities developed and research interests increased, the project has been instrumental in maintaining the statistical capabilities for designing and analyzing tagging studies to meet these expanded objectives. This report describes three alternate methods for constructing confidence intervals for the probability of survival of salmon smolt between hydroelectric facilities. Confidence intervals constructed using a non-parametric bootstrap procedure and intervals constructed using a profile likelihood method are developed and their performance compared to the standard confidence intervals based on normal probability theory.

The statistical analysis was motivated by the continuing need for better and more realistic analytic tools and models for assessing the status of threatened and endangered salmon runs in the Snake and Columbia River system. In response to agency concerns that program SURPH has provided standard errors that are too small resulting in too narrow confidence intervals, we have investigated these properties. We compared standard calculations with the two alternative approaches mentioned above. While normal theory intervals performed adequately, we propose implementation of the profile likelihood method which has the additional desirable property of symmetric errors of over- and under-estimation.

ABSTRACT

Objectives

Confidence intervals for survival probabilities between hydroelectric facilities of migrating juvenile salmonids can be computed from the output of the SURPH software developed at the Center for Quantitative Science at the University of Washington. These intervals have been constructed using the estimate of the survival probability, its associated standard error, and assuming the estimate is normally distributed. In order to test the validity and performance of this procedure, two additional confidence interval procedures for estimating survival probabilities were tested and compared using simulated mark-recapture data. Intervals were constructed using normal probability theory, using a percentile-based empirical bootstrap algorithm, and using the profile likelihood concept. Performance of each method was assessed for a variety of initial conditions (release sizes, survival probabilities, detection probabilities). These initial conditions were chosen to encompass the range of parameter values seen in the 1993 and 1994 Snake River juvenile salmonid survival studies. The comparisons among the three estimation methods included average interval width, interval symmetry, and interval coverage.

Results

We found that the three methods produced nearly identical results in cases where a large number of fish were detected after the initial release, whether this was due to a large release size, a high detection probability, or a high survival probability. In simulations with fewer fish detected after release, the normal theory intervals typically provided nominal coverage and the most narrow width but with a high degree of asymmetry, indicating that the intervals tended to underestimate the survival probability more often than overestimating it. In these cases, the bootstrap percentile and profile likelihood methods also produced intervals with nominal coverage, but with much better symmetry at the expense of slightly wider intervals.

Recommendations

In the simulations, the bootstrap intervals and profile likelihood intervals both provided nominal coverage and had the additional property of symmetry. Due to the high computational cost of the

bootstrap procedure it is recommended that only the profile likelihood procedure be additionally implemented into the SURPH survival analysis software. The choice between normal theory intervals and profile likelihood intervals becomes a choice between the narrower, asymmetric intervals resulting from the normal theory, and the wider, symmetric intervals based on the profile likelihood. If there is equal concern about overestimation and underestimation of parameters, symmetric intervals would be desirable and the profile likelihood method should be used. Thus, our recommendation is that profile likelihood methods be incorporated into SURPH and that the profile likelihood should be the method for constructing confidence intervals for survival probabilities for juvenile salmonids on the Columbia and Snake River systems. For most river conditions, the resulting intervals will be similar to the normal theory intervals providing added assurance of reliable estimates.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF FIGURES (Continued)

TABLE OF TABLES

TABLE OF TABLES (Continued)

Introduction

For several years the survival of out-migrating chinook salmon (*Oncorhynchus tschawyscha*) and steelhead (*O. mykiss*) has been assessed on the Snake River system in Washington by the National Marine Fisheries Service (NMFS) and the University of Washington (e.g. Muir et al. 1995, Iwamoto et al. 1994). These salmon smolt have been individually fitted with PIT-tags that can be detected at slide gate facilities at subsequent downstream hydroelectric facilities. Statistical survival models similar to those developed by Cormack (1964), Jolly (1965), and Seber (1965) and extended by Skalski et al.(1993), Smith (1991), and Hoffman (1993) have been used to estimate survival probabilities and their associated standard errors. These algorithms have been implemented into the computer program SURPH.1 (Smith et al. 1994) developed at the Center for Quantitative Science of the School of Fisheries at the University of Washington.

Recently concern has been voiced that the standard errors reported for the estimates of survival probabilities between hydroelectric projects may be too small, resulting in confidence intervals narrower than appropriate, and thus, with less than nominal (1-α) coverage. To test this contention, we compared two alternative methods of confidence interval estimation to the standard normal theory approach using Monte-Carlo simulations of survival data. The first alternative method used percentile based confidence intervals generated from a non-parametric bootstrap algorithm. The second method used profile likelihood intervals. Coverage rates and confidence interval widths for the three methods were compared to determine the best estimation procedure and discern differences among the methods.

Methods

Monte-Carlo simulation

We started with a fixed release size and a known set of capture and survival probabilities. These initial conditions were used to randomly generate sets of capture histories for R simulated animals, where R represents the release size. We simulated a simple case in which there is an initial release, followed by two subsequent downriver recapture locations. Thus the initial conditions consisted of a release size R; a probability of detection (given that the animal is still alive) at

recapture site 1, denoted by P_1; a probability of surviving release to recapture site 1, denoted by S_1; a probability of detection (given that the animal is still alive) at recapture site 2, denoted by P_2; and a probability of surviving between recapture sites 1 and 2, denoted by S_2 (Fig. 1). To compare this simulated scenario to the primary releases on the Snake River, the initial release would be at the head of Lower Granite Reservoir. Then the downriver recapture sites would be Lower Granite Dam and Little Goose Dam, respectively. For simplicity, we assumed that $S_1 = S_2 = S$ and $P_1 = P_2 = P$.

Release site   Recapture site 1   Recapture site 2

$$S_1 \qquad\qquad\qquad S_2$$

$$R \qquad\qquad\qquad P_1 \qquad\qquad\qquad P_2$$
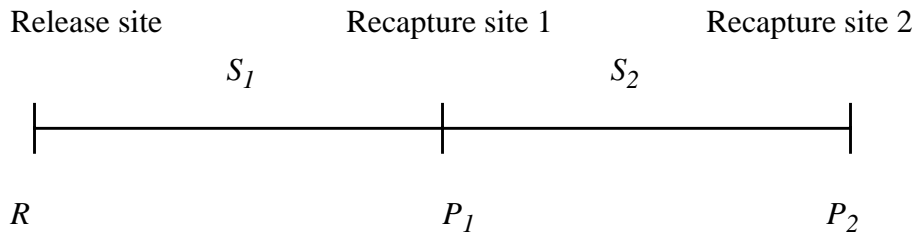
Fig. 1. Mark-recapture scenario used in the simulations with an initial release and two subsequent downriver detections.

The fate of each animal was simulated independently based on draws from independent Bernoulli random variables. Each animal required up to four independent draws, two to determine survival in the intervals, and, if alive, two more to determine detections. For example, a capture history of 110 would indicate that the animal in the initial release (as indicated by the initial 1), was alive and detected at recapture site 1 (as indicated by the second 1) but was not detected at recapture site 2 (as indicated by the 0). If we observed an animal with this capture history, we

would not be able to determine whether that individual had died or had survived but escaped detection at the second site. This simulation process resulted in the creation of a data file consisting of R independent capture histories for the R animals. This entire process was then repeated some large number, N, of times (typically 1000) so that for each set of initial conditions there were N independent sets of capture history data.

Using the Cormack (1964) model there are three parameters which can be estimated from this data: the probability of survival from the initial release to the first recapture site, S_1; the probability of recapture at the first recapture site, P_1; and the joint probability of survival from the first recapture site to the second recapture site and recapture at the second recapture site $\lambda = S_2 P_2$. The two components of this last parameter (i.e. S_2 and P_2) are not separately estimable using the standard Cormack model.

These data files were then analyzed using the Cormack estimators for S_1, P_1, and λ, along with normal theory confidence intervals, non-parametric bootstrap confidence intervals, and profile likelihood confidence intervals. These intervals were computed for the first period survival probability, S_1. These procedures focus on S_1 because the survival probabilities are usually the key parameters of interest, while the others may be viewed as nuisance parameters.

Simulation scheme

Simulations were run based on several scenarios for the parameter values. These scenarios were chosen to approximate the ranges of release sizes and survival and capture probabilities typical for the 1993 and 1994 survival studies on the Snake River. In the baseline simulation (scenario 1), the release size, R, was set to 1000; the survival probabilities, S_1 and S_2, were set at 0.6; and the capture probabilities, P_1 and P_2, were set at 0.5. A "star design" simulation scheme (Fig. 2) was carried out so that in each subsequent simulation, one of the three parameters, S_1, P_1, or R was given a greater or lesser value. This resulted in six more simulation scenarios. The parameter values used in the Monte Carlo simulations are summarized in Table 1. In each scenario, the
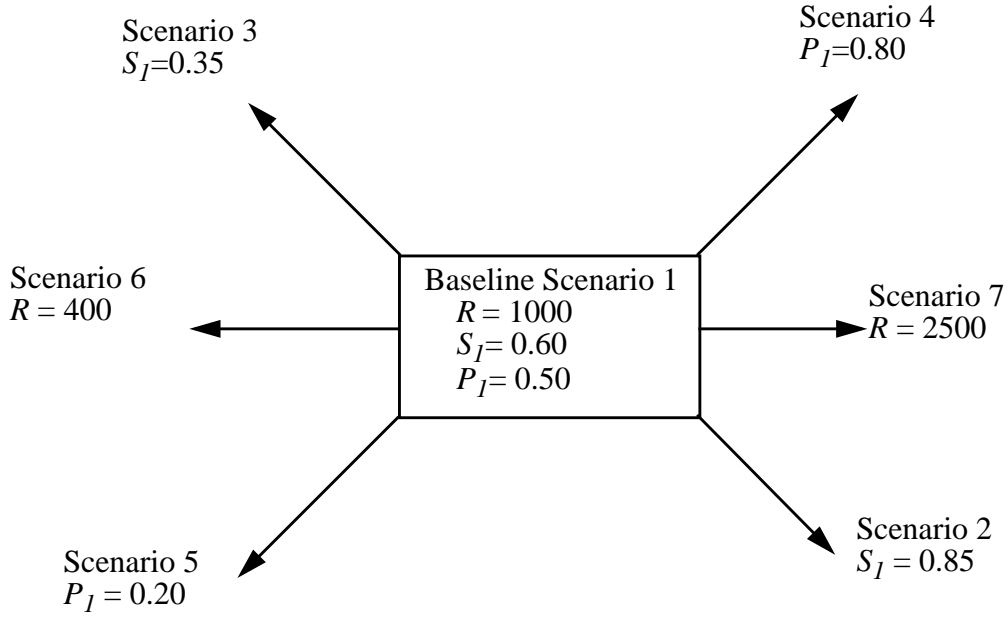
Fig. 2. Schematic of the "star" simulation design. In each of scenarios 2-7 one parameter value is either increased or decreased from its baseline value.

value of λ was set equal to the product of S_1 and P_1. That is, the second period survival and capture probabilities were set equal to the values of the first period survival and capture probabilities. Three additional simulations were performed taking values from 1993 and 1994 Snake River survival studies as the initial conditions. The parameters for these simulations are given in Table 2.

Table 1: Input parameter values for "star" design simulation scenarios.

Scenario number	R	S	P	N	λ
1 (baseline)	1000	0.60	0.50	1000	0.30
2 (high S)	1000	0.85	0.50	1000	0.425
3 (low S)	1000	0.35	0.50	1000	0.175
4 (high P)	1000	0.60	0.80	1000	0.48
5 (low P)	1000	0.60	0.20	1000	0.12
6 (low R)	400	0.60	0.50	1000	0.30
7 (high R)	2500	0.60	0.50	1000	0.30

Table 2: Input parameter values for simulation scenarios based on 1993-1994 Snake River juvenile salmonid survival studies.

Simulation	R	S_1	P_1	S_2	P_2
1993 primary Chinook release 7	1405	0.886	0.531	0.869	0.449
1994 primary Chinook release 4	1190	0.941	0.402	0.723	0.241
1994 primary Chinook release 9	542	0.933	0.307	0.688	0.134

For each of the three methods, 90% and 95% confidence intervals were computed from each of the N data sets. Recall that for each scenario N=1000 data sets were simulated and confidence intervals computed with the three methods. These methods were compared for coverage, average interval width, and interval symmetry. Coverage was measured as the percentage of intervals (out of N) that included the true value of S_1. A procedure for constructing 95% confidence intervals should produce actual coverage close to 95%. Average interval width is considered since we would like the smallest interval that achieves nominal coverage. Symmetry is a desirable, but not crucial, property for the confidence intervals. We define it as the degree to which the confidence interval tends to miss the true value of S_1 equally from the right and from the left. Ideally, a 95% confidence interval would not include the true value of S_1 due to overestimation 2.5% of the time and, likewise, due to underestimation 2.5% of the time. Here, underestimation is taken to mean that the confidence interval lies entirely below (*i.e.* includes only values less than) the true value of the parameter.

Cormack model

In the Cormack likelihood model, closed form estimators can be found for the capture and survival probabilities and for their associated standard errors. If we let n_{100}, n_{101}, n_{110}, and n_{111} represent the number of animals from the R animals released that have the capture history indicated by the subscript, then the multinomial likelihood can be written as:

$$L(S_1, P_1, \lambda | n_{100}, n_{101}, n_{110}, n_{111}) = \binom{R}{n_{100}, n_{101}, n_{110}, n_{111}} \bullet$$

$$(S_1 P_1 \lambda)^{n_{111}} (S_1 P_1 (1 - \lambda))^{n_{110}} (S_1 (1 - P_1) \lambda)^{n_{101}} (1 - S_1 P_1 - S_1 \lambda + S_1 P_1 \lambda)^{n_{100}} .$$

The estimate for the first period survival probability, S_1, can be expressed as:

$$\hat{S}_1 = \frac{(n_{110} + n_{111})(n_{101} + n_{111})}{Rn_{111}}$$

and a closed form expression for $Var(\hat{S}_1 | S_1)$ can be computed using the delta method.

Normal based confidence intervals

Using the Cormack model as given above, $100(1-\alpha)$ % confidence intervals can be computed as:

$$\hat{S}_1 \pm Z_{\frac{\alpha}{2}} \bullet \sqrt{\hat{Var}(\hat{S}_1 | S_1)}$$

where Z_α is the value of the standard normal variable Z such that the area to the right of Z_α is α. Thus a 95% confidence interval for S_1 is given by:

$$\hat{S}_1 \pm 1.96 \bullet \hat{S.E.}(\hat{S}_1 | S_1)$$

where $\hat{S.E.}(\hat{S}_1 | S_1)$ is the standard error of the estimate, equal to the square root of

$$Var(\hat{S}_1 | S_1) .$$

Non-parametric bootstrap percentile confidence intervals

For each data set generated, a bootstrap procedure (Efron and Tibshirani 1993) was used to compute estimates of the survival probability, S_1, and its standard error. Bootstrapping involves resampling from the data (in this case, the individual capture histories) with replacement. If there are R individuals in the original set of capture histories, there will be R "individuals" in the bootstrap data set. Capture histories of some animals in the original data set may appear more than once while others will not appear at all. This resampling procedure is then repeated a large number (usually denoted by B) of times. Values of B from 1000 to 10,000 are typical. A natural 100 $(1-\alpha)$ % confidence interval for S results from taking the 100 $(\alpha/2)$ and 100 $(1-\alpha/2)$ percentiles of the B estimates of S. For example, a 95% confidence interval is formed by taking the 2.5 percentile as the lower limit and the 97.5 percentile as the upper limit. We calculated 90% and 95% bootstrap percentile confidence intervals in this manner.

Profile likelihood confidence intervals

The profile likelihood concept has been gaining favor in the biometrical community (Cormack 1992; Hirst 1994) and was also used here to generate confidence intervals for S_1. This method is based on the likelihood ratio test (Kalbfleish and Sprott 1970). In order to construct a profile confidence interval for S_1, we first found the likelihood value for the Cormack likelihood model evaluated at the maximum likelihood estimates (MLE), $\hat{S}_1$, $\hat{P}_1$, and $\hat{\lambda}$. We denote this likelihood by $L(\hat{S}, \hat{P}, \hat{\lambda})$. The profile likelihood confidence interval for S_1 can then be expressed by the equation:

$$CI(S_1) = \left\{ S_1: -2\ln\left(\frac{L(S_1, \hat{P}_1', \hat{\lambda}')}{L(\hat{S}_1, \hat{P}_1, \hat{\lambda})}\right) \le \chi^2_{1, 1-\alpha} \right\}$$

where $L(S_1, \hat{P}_1', \hat{\lambda}')$ is the value of the likelihood function evaluated at some S_1 other than the MLE and at the $\hat{P}'$ and $\hat{\lambda}'$ that maximize the likelihood given this value for S_1. The confidence interval estimates are based on the asymptotic χ^2 distribution of the log-likelihood ratio given

7

above. Computationally, we find the confidence interval using a simple direct search routine beginning at the MLE, then choosing values of S_1 above and below the MLE until the above inequality is no longer satisfied. Those values of S_1 that first reject the inequality become the upper and lower endpoints of the confidence interval based on the profile likelihood. Again, 90% and 95% confidence intervals for S_1 were computed in this manner.

Results

Confidence interval comparison

Results for the baseline simulation, Scenario 1, are presented in Table 3. For both the 90% and 95% confidence intervals, each method produced close to nominal coverage and similar widths. The bootstrap percentile and profile likelihood interval were somewhat more symmetric. Figure 3 provides a plot of coverages for different values of S_1 for the three different confidence interval methods, based on the 95% intervals. We would expect maximum coverage at the true value of the parameter, $S_1 = 0.6$, which is what is seen in the cases of the profile likelihood and bootstrap intervals. Maximum coverage for the normal theory intervals occurred at a point just below 0.60 (0.58) indicating a tendency for the normal theory intervals to slightly underestimate the survival probability.

Table 3: Simulation results for Scenario 1 (Baseline).

Confidence coefficient	Method	Actual coverage	Average width	% under-estimated	% over-estimated
90%	Normal theory	89.8	0.135	6.1	4.1
90%	Bootstrap percentile	89.3	0.137	5.1	5.6
90%	Profile likelihood	89.5	0.136	5.0	5.5
95%	Normal theory	94.1	0.161	4.4	1.5
95%	Bootstrap percentile	94.0	0.164	3.0	3.0
95%	Profile likelihood	94.0	0.163	2.8	3.2

Fig. 3. Percent coverage of 95% confidence intervals for three alternative methods of interval construction at different values of S_1, when S_1=0.60 (vertical line) in the Monte-Carlo simulation (Scenario 1).

Table 4 contains the simulation results for Scenario 2, where S_1 has been increased above the baseline to 0.85. Here the three methods performed nearly identically. The normal theory intervals had slightly larger than nominal coverage and comparable width to the bootstrap percentile and profile likelihood confidence intervals. The normal theory intervals were slightly less symmetric than the other two methods. The 95% coverage plots (Fig. 4) again coincided for the bootstrap percentile and profile likelihood intervals, whereas the normal theory intervals tended to slightly underestimate the survival probability, with maximum coverage occurring at a value slightly less than 0.85.

Table 4: Simulation results for Scenario 2 (High S).

Confidence coefficient	Method	Actual coverage	Average width	% under-estimated	% over-estimated
90%	Normal theory	90.8	0.119	5.2	4.0
90%	Bootstrap percentile	90.0	0.119	4.7	5.3
90%	Profile likelihood	90.3	0.120	4.4	5.3
95%	Normal theory	95.2	0.142	3.2	1.6
95%	Bootstrap percentile	94.3	0.143	2.8	2.9
95%	Profile likelihood	94.7	0.143	2.8	2.5

Table 5 contains the simulation results for Scenario 3, where S_1 has been reduced below the baseline level to 0.35. This results in fewer animals being seen at the recapture sites and hence decreased precision in the estimate of the survival probability. Note that all three methods still produced intervals of nominal coverage but that the intervals on average were substantially wider than in the two previous scenarios. The normal intervals were quite asymmetric in this simulation, as is especially apparent in a comparison of the 90% confidence intervals for S. Again, the 95% coverage plot (Fig. 5) showed agreement between the bootstrap and profile likelihood intervals while the normal theory intervals had maximum coverage at a value less than 0.35 (0.33).

Fig. 4. Percent coverage of 95% confidence intervals for three alternative methods of interval construction at different values of S_1, when S_1=0.85 (vertical line) in the Monte-Carlo simulation (Scenario 2).

Fig. 5. Percent coverage of 95% confidence intervals for three alternative methods of interval construction at different values of S_1, when S_1=0.35 (vertical line) in the Monte-Carlo simulation (Scenario 3).
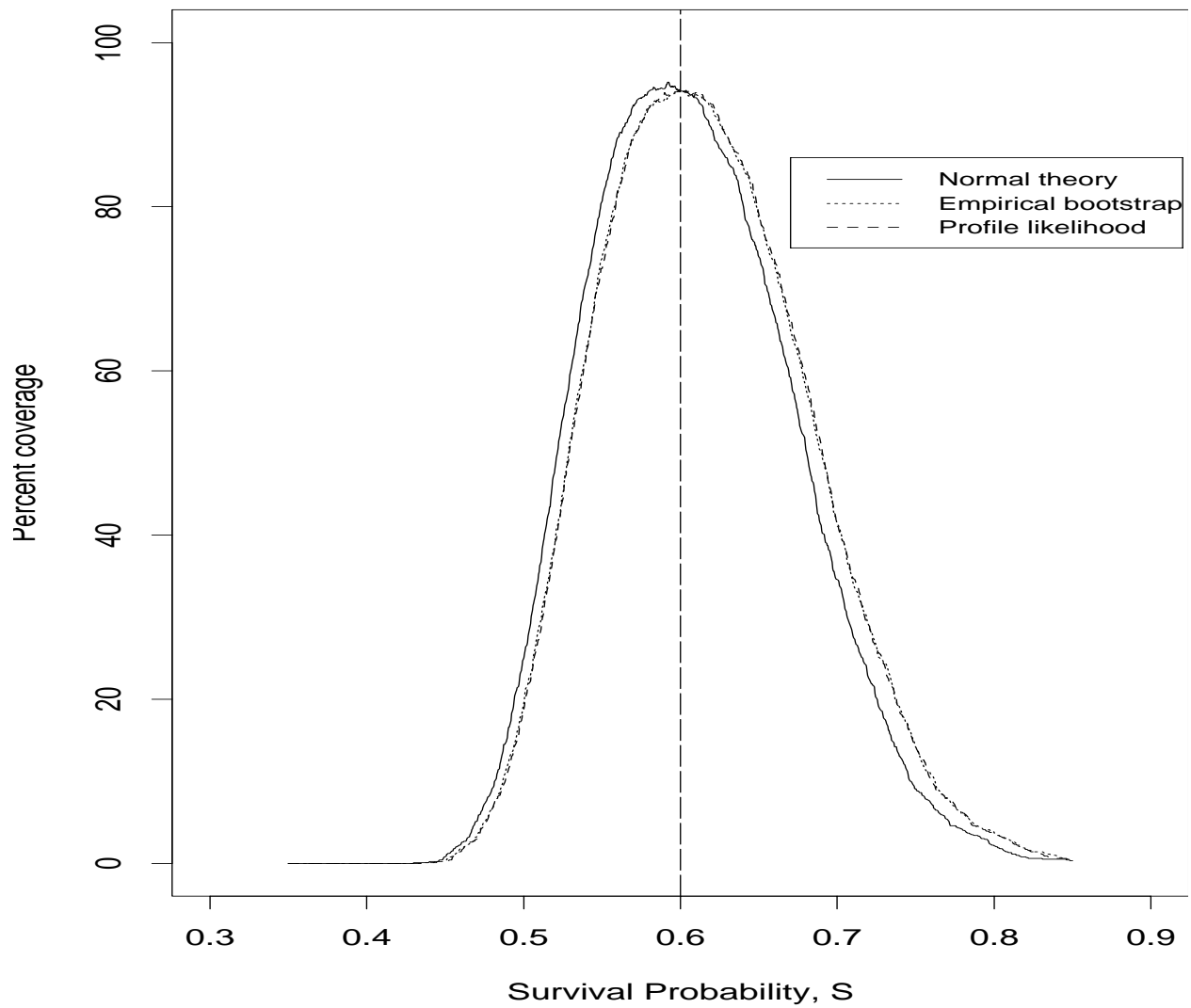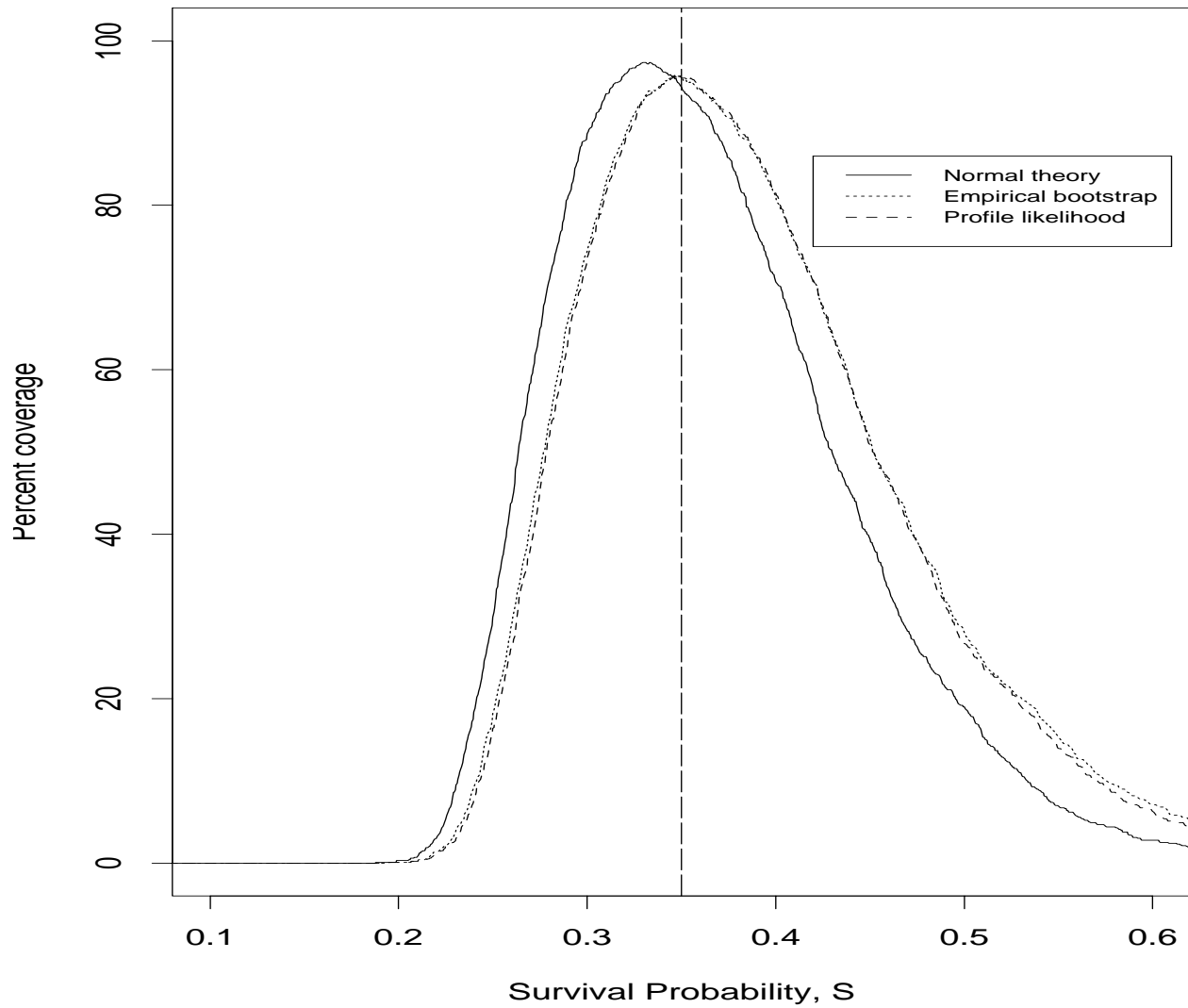
Table 5: Simulation results for Scenario 3 (Low S).

Confidence coefficient	Method	Actual coverage	Average width	% under-estimated	% over-estimated
90%	Normal theory	89.6	0.148	8.1	2.3
90%	Bootstrap percentile	89.8	0.156	5.1	5.1
90%	Profile likelihood	89.7	0.154	5.0	5.3
95%	Normal theory	94.4	0.177	5.0	0.6
95%	Bootstrap percentile	95.6	0.189	1.4	3.0
95%	Profile likelihood	95.5	0.186	1.5	3.0

Table 6 contains the simulation results for Scenario 4, where P_1, the probability of detection at the first recapture site has been increased above the baseline level to 0.80. Again, this was a scenario where many animals were seen on multiple occasions so that there was sufficient information to estimate S_1 with good precision. The three methods produced nearly identical results in terms of coverage, width, and symmetry. The average interval was quite narrow due to the abundant information and the resulting high degree of precision. The three methods also coincided on the 95% coverage plots (Fig. 6), with all three indicating maximum coverage at $\hat{S}_1 = 0.60$, and coverage dropping off sharply for values of $\hat{S}_1$ below and above 0.60.

Table 6: Simulation results for Scenario 4 (High P).

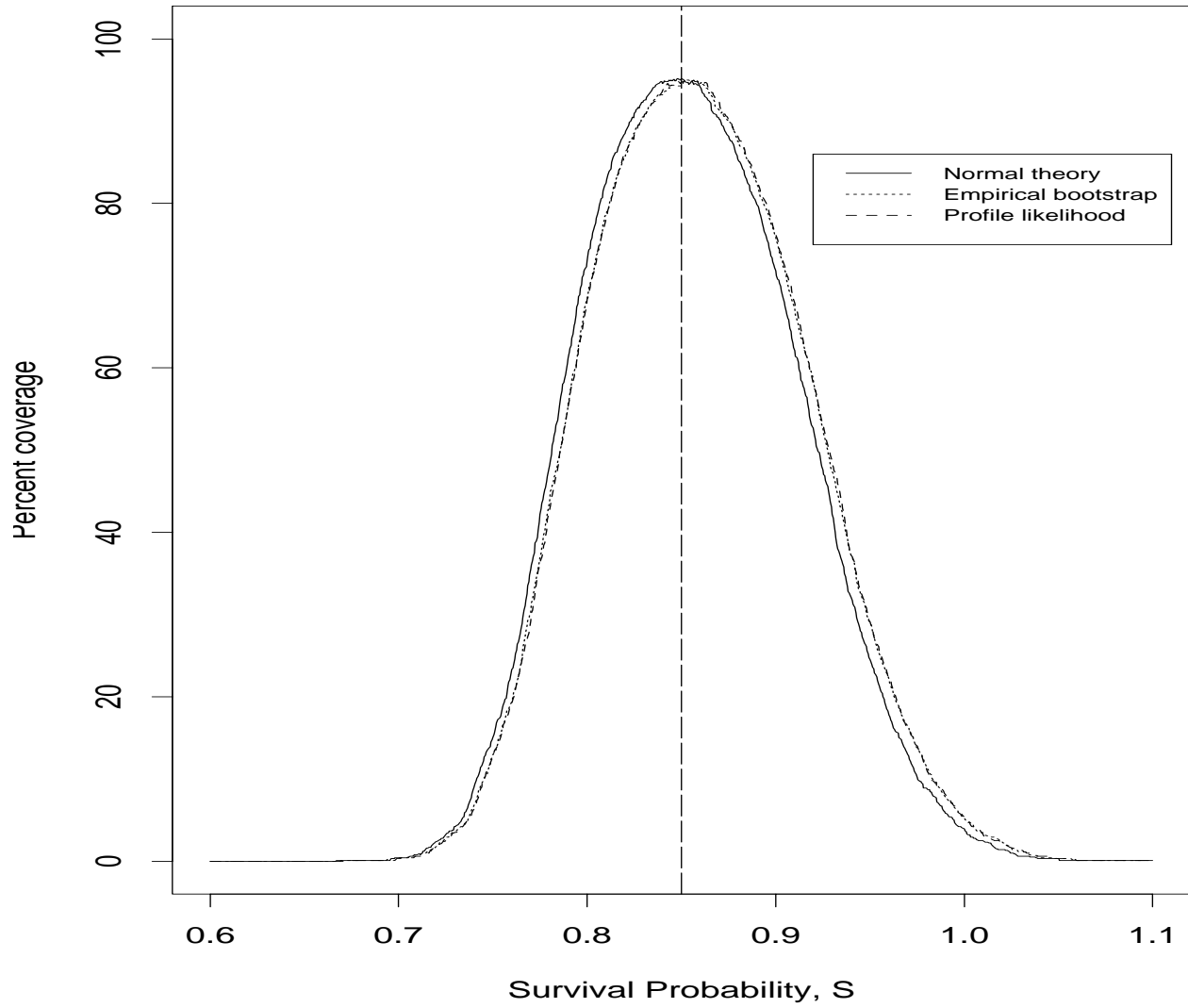| Confidence coefficient	Method	Actual coverage	Average width	% under-estimated	% over-estimated
90%	Normal theory	89.6	0.0661	5.6	4.8
90%	Bootstrap percentile	89.6	0.0661	5.4	5.0
90%	Profile likelihood	89.6	0.0665	5.3	5.1
95%	Normal theory	94.7	0.0788	3.1	2.2
95%	Bootstrap percentile	94.4	0.0789	2.9	2.7
95%	Profile likelihood	95.1	0.0792	2.5	2.4

Fig. 6. Percent coverage of 95% confidence intervals for three alternative methods of interval construction at different values of S_1, when $S_1=0.60$ (vertical line) in the Monte-Carlo simulation (Scenario 4).
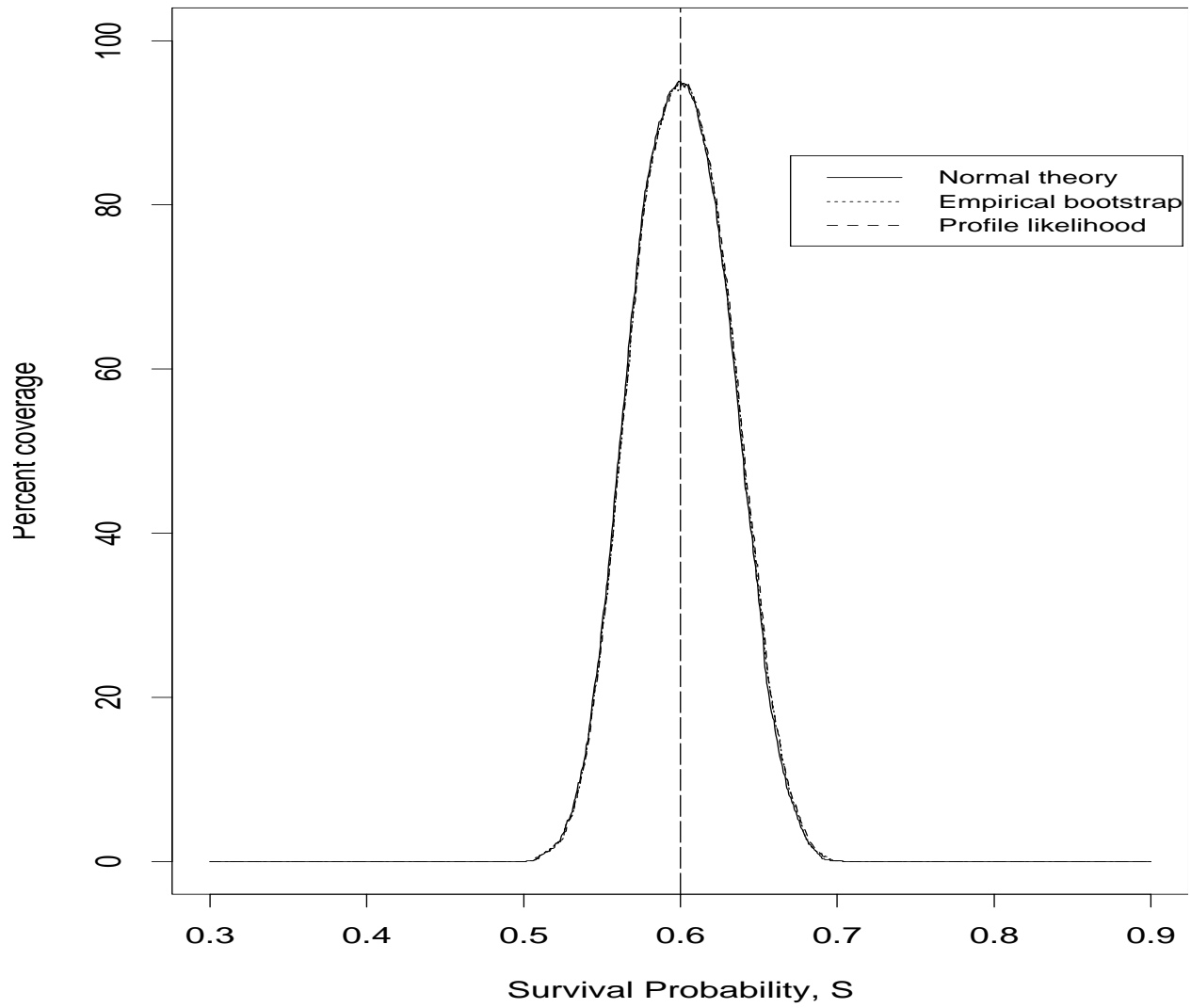
Table 7 contains the simulation results for Scenario 5, where P_1, the probability of detection at the first recapture site has been reduced from the baseline level to 0.20. This reflects another scenario where few animals are detected subsequent to the initial release, resulting in estimates with low precision. The normal theory intervals had the largest coverage for the 90% intervals, yet the lowest for the 95% intervals. Also, the normal theory intervals were highly asymmetric, tending to underestimate the value of the survival probability. In this case, the profile likelihood intervals seemed to perform better overall, with nominal coverage, symmetry, and only slightly wider intervals. The intervals for the bootstrap percentile method were substantially wider. The 95% coverage plot (Fig. 7) shows similar results for the bootstrap and profile likelihood intervals, both achieving maximum coverage at the simulated value, $\hat{S}_1 = 0.60$. Although the normal intervals provide near-nominal coverage at $\hat{S}_1 = 0.60$, the maximum coverage occurs at a value well below 0.60 at 0.47, indicating a tendency to underestimate the true survival probability.

Table 8 contains the simulation results for Scenario 6, where R, the initial release size, has been decreased from the baseline level of 1000 individuals to 400. This reflects another scenario where few animals are detected subsequent to the initial release, resulting in estimates with lower precision. The normal theory intervals had the largest coverage for the 90% intervals and the low-

Table 7: Simulation results for Scenario 5 (Low P).

Confidence coefficient	Method	Actual coverage	Average width	% under-estimated	% over-estimated
90%	Normal theory	91.1	0.504	8.7	0.2
90%	Bootstrap percentile	88.4	0.601	4.5	7.1
90%	Profile likelihood	88.4	0.508	5.9	5.7
95%	Normal theory	93.3	0.601	6.7	0.0
95%	Bootstrap percentile	94.4	0.774	1.7	3.9
95%	Profile likelihood	95.1	0.650	2.5	2.4

Fig. 7. Percent coverage of 95% confidence intervals for three alternative methods of interval construction at different values of S_1, when S_1=0.60 (vertical line) in the Monte-Carlo simulation (Scenario 5).
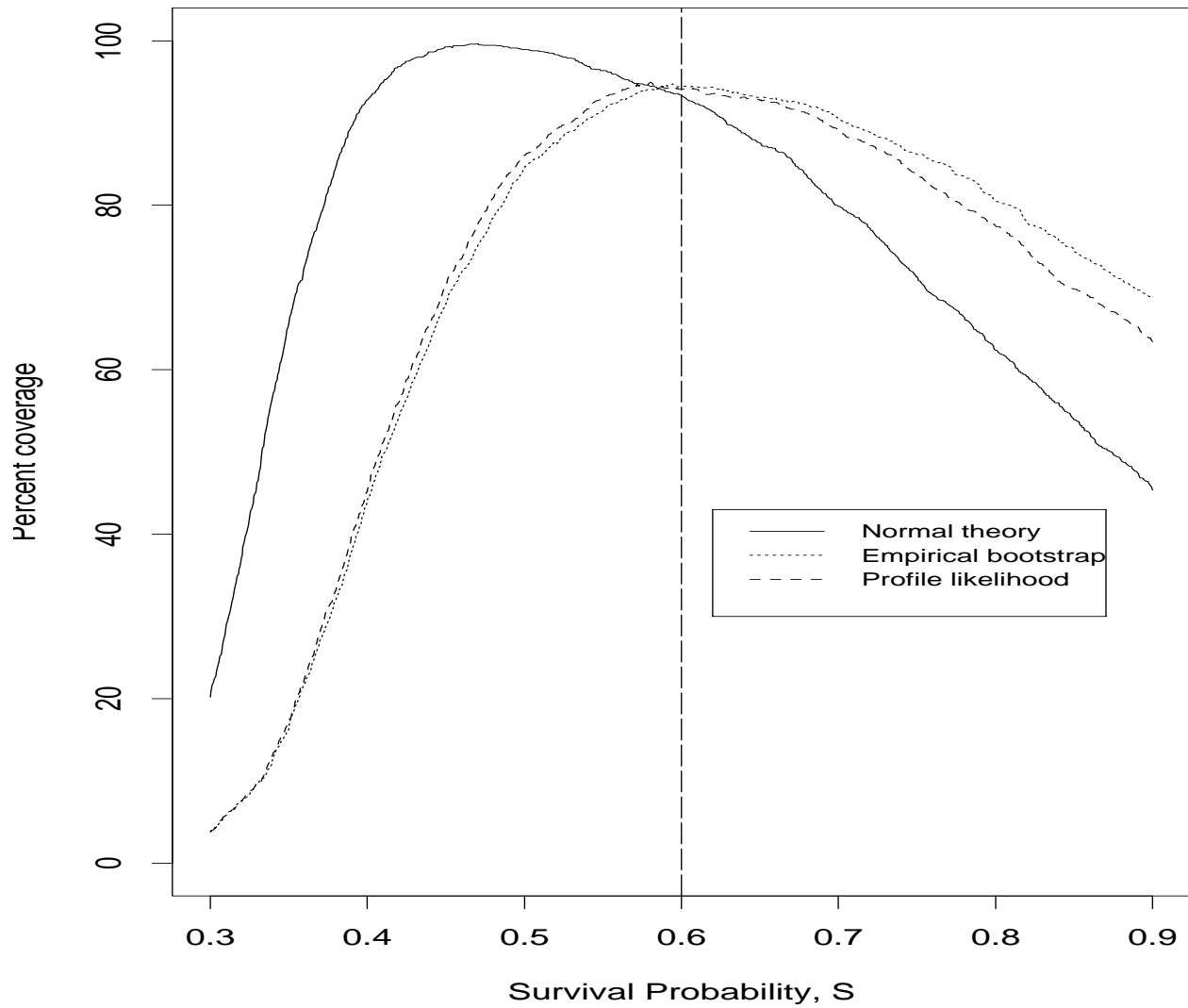
est for the 95% intervals, but all three methods provided essentially nominal coverage. The interval widths were similar, with the normal theory intervals again being slightly narrower, and somewhat more asymmetric. The 95% coverage plot (Fig. 8) shows maximum coverage of the bootstrap and profile likelihood intervals occurred at the simulated value, $S_1 = 0.60$, while the maximum coverage for the normal theory intervals occurred at 0.58, again indicating a tendency for these intervals to underestimate the survival probability.

Table 8: Simulation results for Scenario 6 (Small R).

Confidence coefficient	Method	Actual coverage	Average width	% under-estimated	% over-estimated
90%	Normal theory	90.0	0.217	7.8	2.2
90%	Bootstrap percentile	89.5	0.225	5.9	4.6
90%	Profile likelihood	89.7	0.223	5.6	4.7
95%	Normal theory	94.3	0.258	5.3	0.4
95%	Bootstrap percentile	94.7	0.272	2.8	2.5
95%	Profile likelihood	95.1	0.269	2.7	2.2

Table 9 contains the simulation results for Scenario 7, where R, the initial release size, has been increased from the baseline level of 1000 individuals to 2500. This results in more animals being detected subsequent to the initial release, resulting in estimates with greater precision. Coverages for the three methods were very similar, but somewhat below nominal coverage. The interval widths were nearly identical, and again the normal theory intervals exhibited somewhat greater asymmetry. The 95% coverage plot (Fig. 9) shows maximum coverage for all three methods occurred close to the simulated value, $S_1 = 0.60$, and, as in Scenario 4, coverage dropped off sharply for values of $\hat{S}_1$ below and above 0.60.

Fig. 8. Percent coverage of 95% confidence intervals for three alternative methods of interval construction at different values of S_1, when S_1=0.60 (vertical line) in the Monte-Carlo simulation (Scenario 6).

Fig. 9. Percent coverage of 95% confidence intervals for three alternative methods of interval construction at different values of S_1, when S_1=0.60 (vertical line) in the Monte-Carlo simulation (Scenario 7).

Table 9: Simulation results for Scenario 7 (Large R).

Confidence coefficient	Method	Actual coverage	Average width	% under-estimated	% over-estimated
90%	Normal theory	87.7	0.0846	7.1	5.2
90%	Bootstrap percentile	87.5	0.0850	6.3	6.2
90%	Profile likelihood	87.3	0.0852	6.1	6.6
95%	Normal theory	94.1	0.101	4.0	1.9
95%	Bootstrap percentile	93.8	0.101	3.1	3.1
95%	Profile likelihood	94.2	0.102	2.8	3.0

Additional simulations were run based on conditions observed in the Snake River in 1993 and 1994 (Table 2). The 1993 Snake River primary release 7 (Iwamoto et al. 1994) resulted in parameter estimates indicating high survival from the release in Lower Granite Reservoir to the tailrace of Lower Granite Dam (S_1 = 0.886) and from Lower Granite Dam to Little Goose Dam (0.869), along with reasonably high detection probabilities (0.531, 0.449) at the two dams, respectively. The initial release, R, size was 1405. These parameter values correspond most closely to Scenario 2 and the results of a simulation using these initial conditions confirms this (Table 10). The three methods of confidence interval construction perform nearly identically. The 95% coverage plot also (Fig. 10) confirms this with maximum coverage for the three methods occurring close to

$\hat{S}_1$ = 0.886.

In 1994, detection probabilities on the river varied widely, in large part due to increased spill over the dams late in the season (Muir et al. 1995). The 1994 Snake River Chinook primary release 4 resulted in high estimates of survival probabilities (0.941, 0.723) but low estimates of detection probabilities (0.402, 0.241), along with a release size, R, equal to 1190. This represents
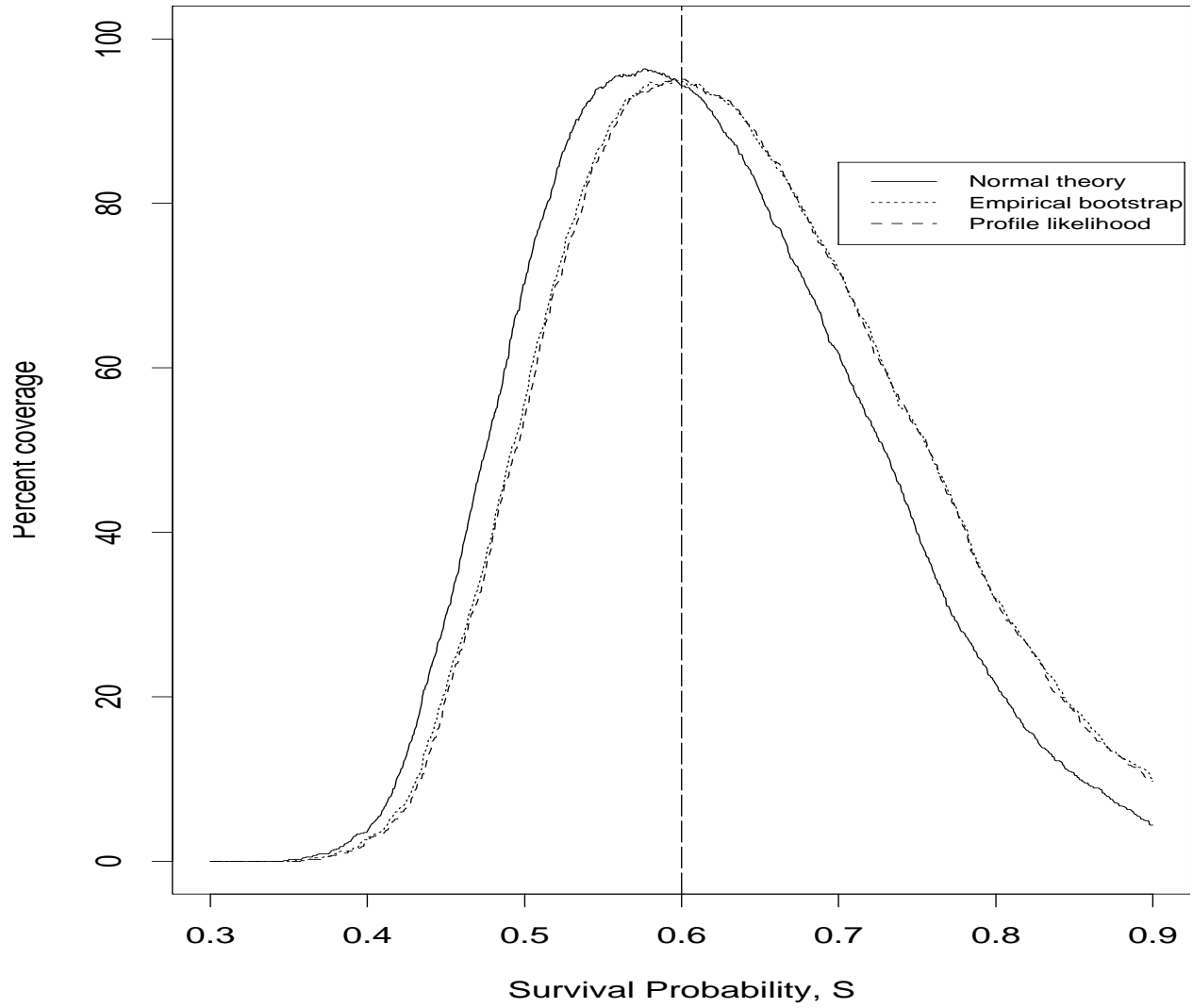
Fig. 10. Percent coverage of 95% confidence intervals for three alternative methods of interval construction at different values of S_1, when S_1=0.886 (vertical line) in the Monte-Carlo simulation (Scenario based on 1993 Snake River chinook primary release group 7).
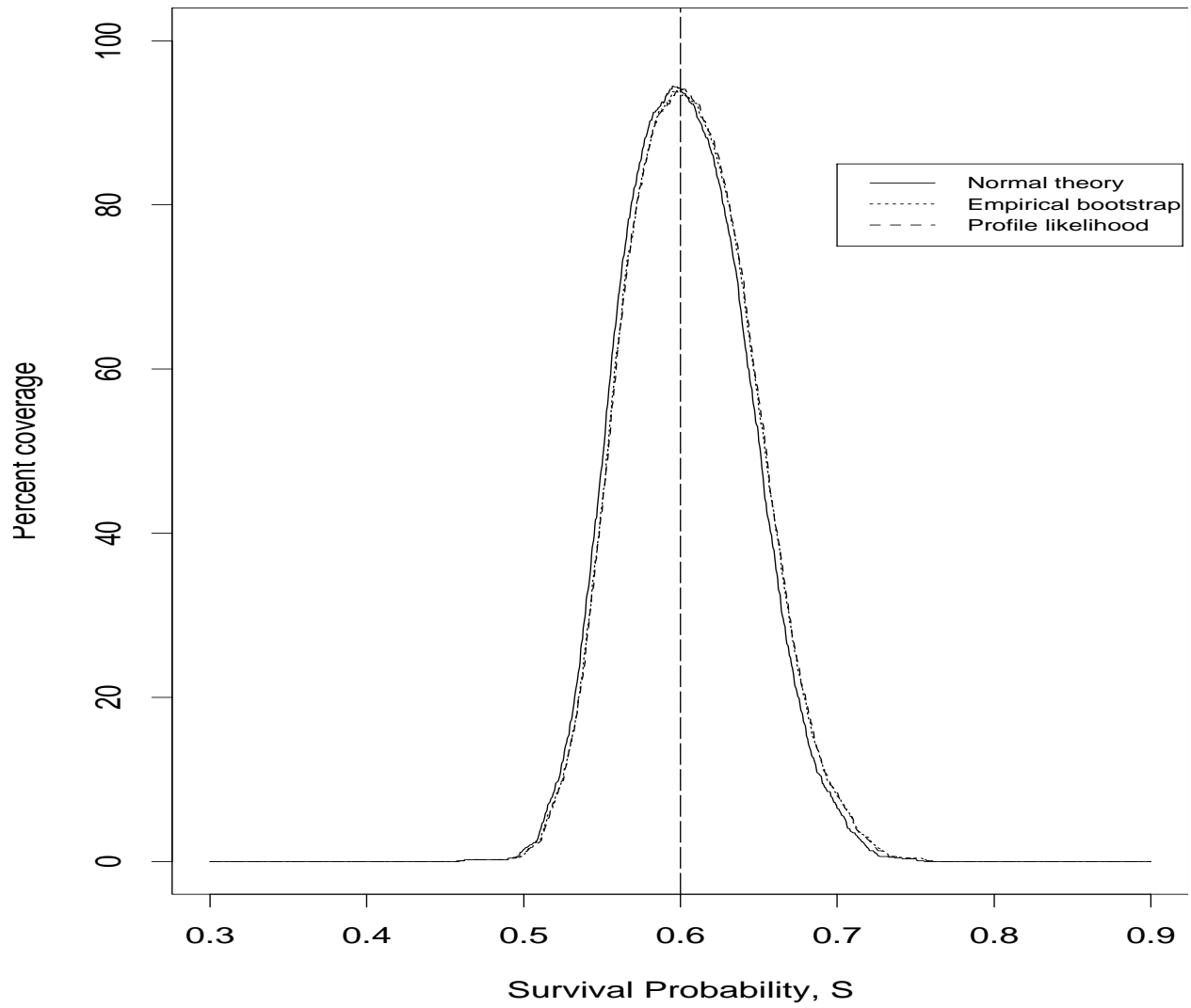
Table 10: Simulation results for 1993 Snake River Chinook primary release 7.

Confidence coefficient	Method	Actual coverage	Average width	% under-estimated	% over-estimated
90%	Normal theory	89.4	0.101	6.8	3.8
90%	Bootstrap percentile	88.2	0.102	6.7	5.1
90%	Profile likelihood	89.2	0.102	6.0	4.8
95%	Normal theory	94.8	0.120	3.8	1.4
95%	Bootstrap percentile	94.4	0.120	3.5	2.1
95%	Profile likelihood	95.1	0.121	3.0	1.9

survival probabilities higher than the baseline scenario, but detection probabilities lower than the baseline. In this simulation (Table 11), the performance of the three methods is comparable. All provide nearly nominal coverage. The normal intervals are again slightly narrower and more asymmetric than the bootstrap and profile likelihood intervals. The bootstrap and profile likelihood intervals have maximum coverage approximately at the simulated value for the survival probability (0.941) (Fig. 11), while the normal theory intervals again have maximum coverage at a point slightly less than the simulated value (0.92).

Table 11: Simulation results for 1994 Snake River Chinook primary release 4.

Confidence coefficient	Method	Actual coverage	Average width	% under-estimated	% over-estimated
90%	Normal theory	89.9	0.250	7.2	2.9
90%	Bootstrap percentile	89.9	0.255	5.5	4.6
90%	Profile likelihood	89.8	0.254	5.5	4.7
95%	Normal theory	94.9	0.298	4.5	0.6
95%	Bootstrap percentile	94.5	0.306	2.8	2.7
95%	Profile likelihood	94.8	0.304	2.5	2.7

Fig. 11. Percent coverage of 95% confidence intervals for three alternative methods of interval construction at different values of S_1, when S_1=0.941 (vertical line) in the Monte-Carlo simulation (Scenario based on 1994 Snake River chinook primary release group 4).
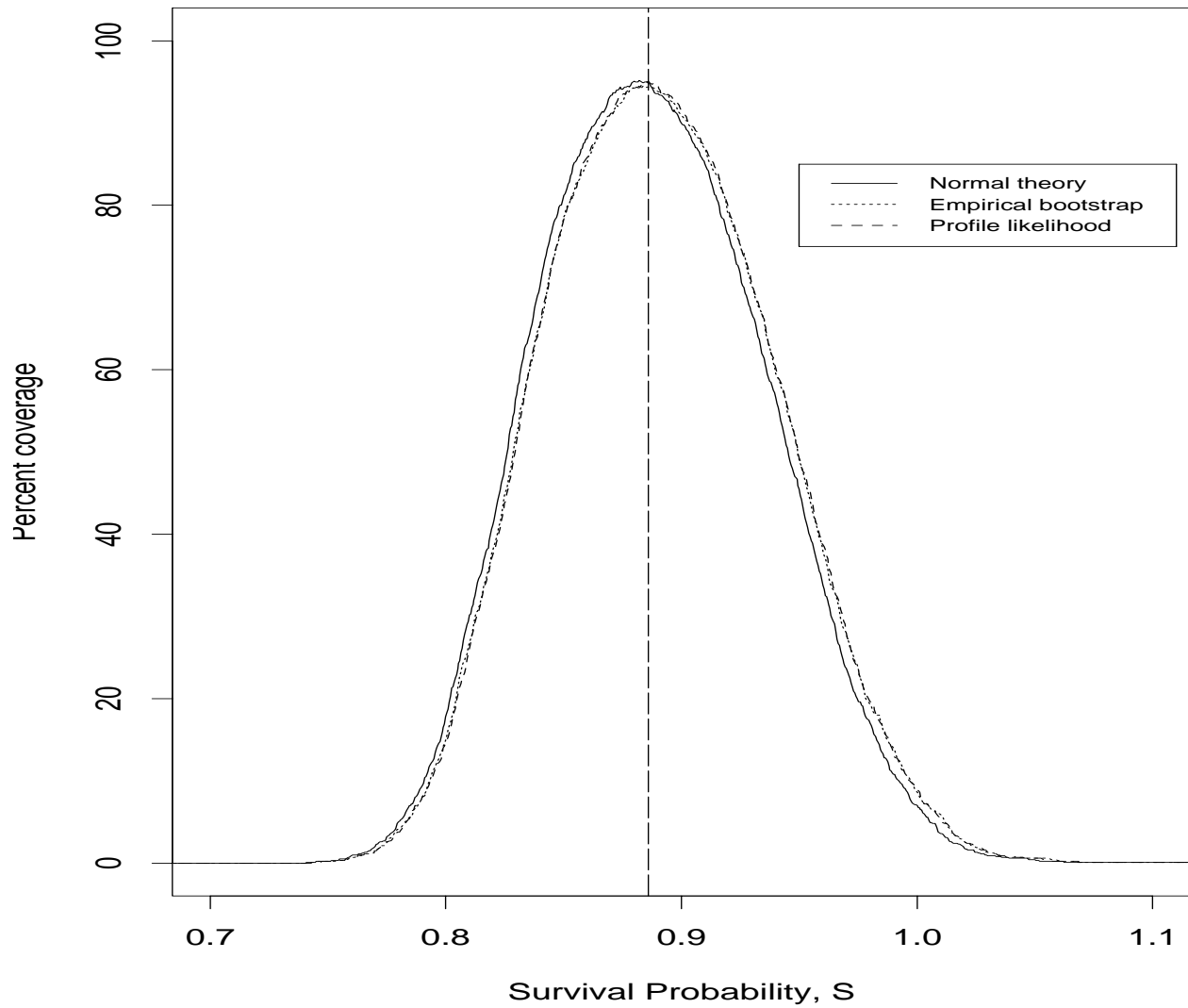
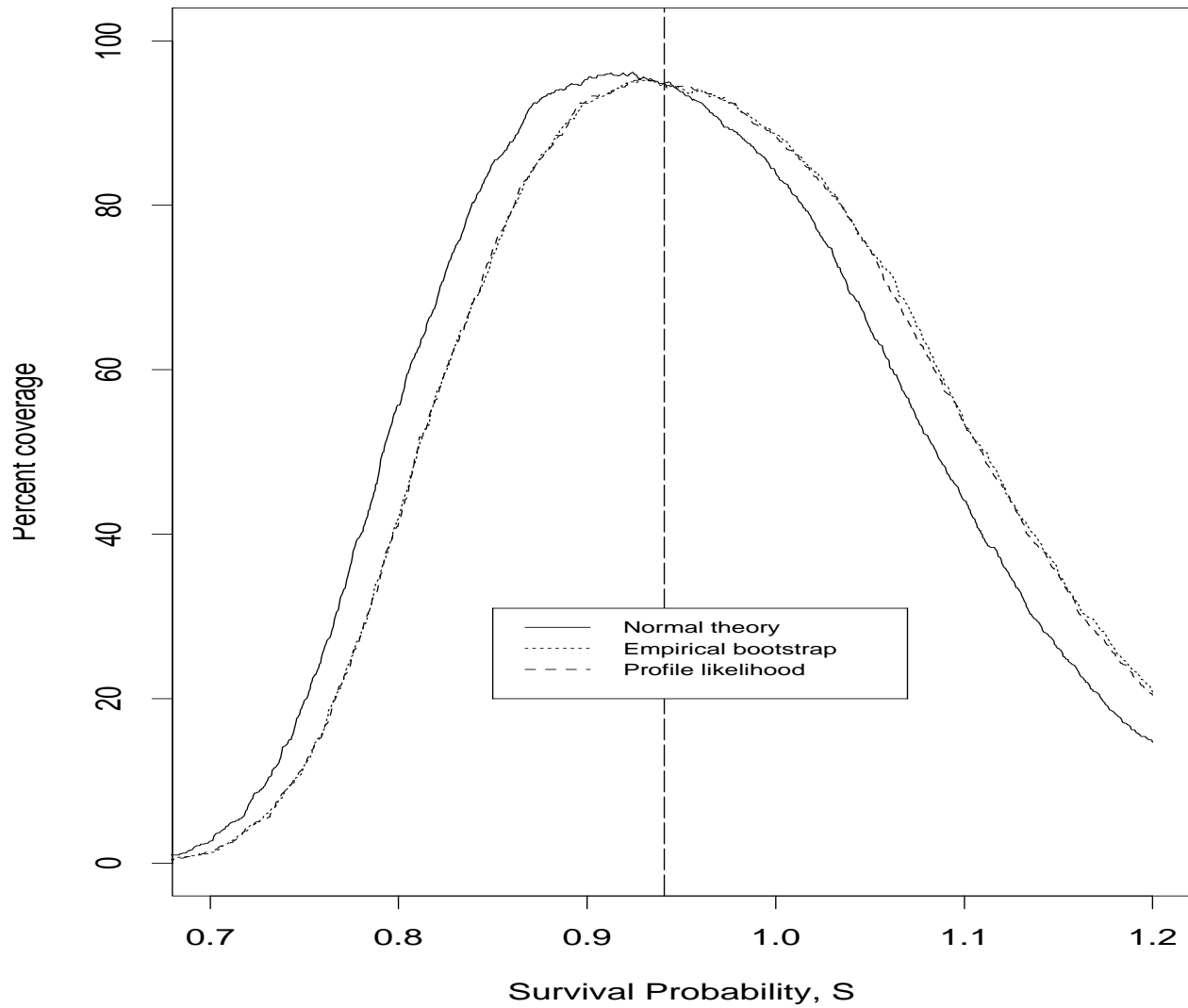The 1994 Snake River Chinook primary release 9 (Muir et al. 1995) resulted in moderately high estimates of survival probabilities (0.933, 0.688) but very low estimates for the detection probabilities (0.307, 0.134), along with a very small release size, R, equal to 542. This represents survival probabilities higher than the baseline scenario, but detection probabilities much lower than the baseline, and a release size much lower than the baseline. This results in poor precision for all three methods of confidence interval construction (Table 12), as evidenced by wide intervals. The coverages for the methods are close to nominal but the widths are so large as to be of little practical value. However, the profile likelihood method results in the narrowest 90% intervals, while normal theory results in the narrowest 95% intervals. Normal theory intervals again are highly asymmetric with a tendency to underestimate the survival probability. The 95% coverage plot (Fig. 12) demonstrates the imprecise nature of the intervals under these initial conditions. Maximum coverage for the normal theory intervals occurs at a value much less than the simulated value of 0.933 (0.75). The bootstrap and profile likelihood intervals, however, do exhibit maximum coverage near (0.95) the simulated value for the survival probability.

Table 12: Simulation results for 1994 Snake River Chinook primary release 9.

Confidence coefficient	Method	Actual coverage	Average width	% under-estimated	% over-estimated
90%	Normal theory	91.6	0.724	8.4	0.0
90%	Bootstrap percentile	91.2	0.856	3.3	5.5
90%	Profile likelihood	91.5	0.659	3.9	4.6
95%	Normal theory	94.8	0.863	5.2	0.0
95%	Bootstrap percentile	95.7	1.096	1.7	2.6
95%	Profile likelihood	95.6	0.887	2.1	2.3

Fig. 12. Percent coverage of 95% confidence intervals for three alternative methods of interval construction at different values of S_I, when S_I=0.933 (vertical line) in the Monte-Carlo simulation (Scenario based on 1994 Snake River chinook primary release group 9).
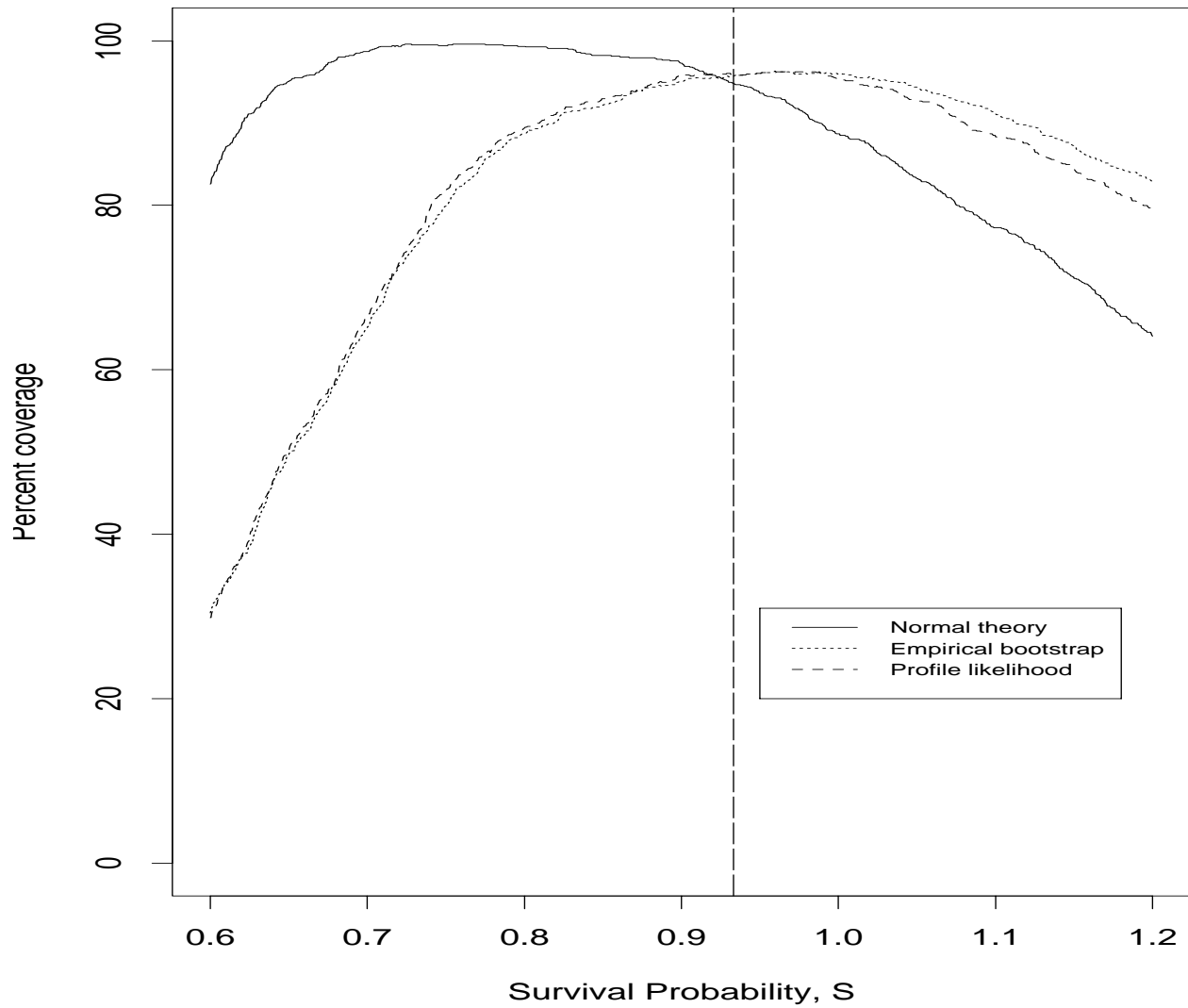
Table 13: Average estimates of the survival probability, S_1, and estimates of variance for the Cormack model and the bootstrap method.

	S_1	$\overline{\hat{S}_1}$	avg. bootstrap $\hat{S}_1$	$Var(\hat{S}_1)$	$\overline{Var(\hat{S}_1 \mid S_1)}$	average bootstrap variance
Scenario 1	0.60	0.603	0.605	0.00172	0.00171	0.00178
Scenario 2	0.85	0.853	0.854	0.00129	0.00132	0.00134
Scenario 3	0.35	0.354	0.360	0.00233	0.00218	0.00257
Scenario 4	0.60	0.600	0.600	0.000404	0.000405	0.000406
Scenario 5	0.60	0.636	0.684	0.0276	0.0286	0.0649
Scenario 6	0.60	0.605	0.611	0.00452	0.00454	0.00510
Scenario 7	0.60	0.601	0.602	0.000696	0.000666	0.000675
1993 - SRPR 7	0.886	0.886	0.887	0.000991	0.000949	0.000963
1994 - SRPR 4	0.941	0.945	0.951	0.00561	0.00592	0.00628
1994 - SRPR 9	0.933	0.975	1.040	0.0512	0.0582	0.131

Parameter estimates and variance comparison

Estimates of the first period survival probability from the Cormack model and the bootstrap algorithm and their associated variances were also calculated (Table 13). Here, $\overline{\hat{S}_1}$ was the average estimate of S_1 from the Cormack model averaged over the N iterations (simulated mark-recapture data sets). The average bootstrap estimate of S_1 was based on the same N data sets. In most cases the average bootstrap estimate was slightly larger than the average Cormack estimate, with both slightly larger than the true value of S_1. In simulations where a large number of individuals were recaptured, the differences were quite small (e.g. Scenarios 1,2,4). When fewer individuals were recaptured, the bias in the methods became larger (e.g. Scenario 5, 1994 Snake River chinook primary release 9), particularly for the bootstrap estimates, resulting in estimates of S_1 greater than the true value.

In Table 13, $Var(\hat{S}_1)$ is the empirical variance of the Cormack estimates defined as

$$Var(\hat{S}_1) = \frac{\sum_{i=1}^{N} (\hat{S}_{1,ij} - \overline{\hat{S}_1})^2}{N-1} \ .$$

The empirical variance is our best estimate of the true sampling variance of the release-recapture method. Whereas $\overline{Var(\hat{S}_1|S_1)}$ gives the average of the Cormack model-based estimates of the variance of S_1, so that

$$\overline{Var(\hat{S}_1|S_1)} = \frac{\sum_{i=1}^{N} Var(\hat{S}_{1,i}|S_1)}{N} \ .$$

The average bootstrap variance is calculated as the mean of the empirical variance of the bootstrap estimates. That is,

$$\text{Average bootstrap variance} = \frac{\sum_{i=1}^{N} \left[\frac{\sum_{j=1}^{B} (\hat{S}_{1,ij} - \overline{\hat{S}_{1,i}})^2}{B-1} \right]}{N}$$

where B is the number of bootstrap iterations, and $\hat{S}_{1,ij}$ represents an individual estimate from the bootstrap procedure.

Again, in the scenarios where a large number of individuals are recaptured, the average model-based variance, $\overline{Var(\hat{S}_1|S_1)}$, and the average bootstrap variance produce comparable estimates of the true variance of the parameter, S_1. In scenarios where fewer individuals are recaptured, the average model-based variance produced estimates of the variance close to the estimate of the true variance (i.e. the empirical variance) while the bootstrap variance tended to overestimate the sampling variance.

Conclusions

Based on Monte-Carlo simulations and an underlying Cormack model, normal theory confidence intervals for survival probabilities provide close to nominal coverage for the range of release sizes and survival and detection probabilities typically seen on the Snake River. Alternative methods of confidence interval construction based on a bootstrap algorithm and based on the profile likelihood were also calculated. These intervals also provided close to nominal coverage usually at the expense of slightly greater width but with the benefit of added symmetry. In cases of large release sizes, and high survival and/or high detection probabilities the three methods produced virtually indistinguishable results. When these conditions were not met, the three methods each continued to provide near nominal coverage but differences in width and symmetry became more apparent. The greatest discrepancies among the methods were seen in the simulation based on the 1994 Snake River Chinook primary release 9. In this simulation, differences between methods were exhibited in interval width and symmetry. The simulations that resulted in the narrowest normal theory confidence intervals (i.e. based on large release size and high survival and/or detection probabilities) also showed the greatest agreement among the three methods of confidence interval construction. In all cases, bootstrap percentile intervals and the profile likelihood based intervals produced very similar results.

In cases where the three methods do not agree, the choice of the appropriate method of confidence interval procedure will depend on the desired properties. Based on the results presented here, the bootstrap percentile intervals would not be the appropriate choice in any of the scenarios. The intervals produced by the bootstrap algorithm were never superior to the other methods and the bootstrap method is the most computationally intensive.

Normal theory and the profile likelihood both produced intervals of nominal coverage under a wide range of simulated conditions. In the cases where they produced different results, a choice between these methods is a choice between the narrower, asymmetric intervals resulting from the normal theory, and the wider, symmetric intervals based on the profile likelihood. In a situation where there is equal concern about overestimation and underestimation of parameters, symmetric intervals would be desirable and the profile likelihood method should be used. Based on this, our recommendation is that profile likelihood methods be incorporated into SURPH and that the profile likelihood should be the method for constructing confidence intervals for survival probabili-

ties for juvenile salmonids on the Columbia and Snake River systems. For most river conditions, the resulting intervals will be similar to the normal theory intervals providing added assurance of reliable estimates.

Literature cited

Cormack, R.M. 1964. Estimates of survival from the sightings of marked animals. Biometrika 51:429-438.

Cormack, R.M. 1992. Interval estimation for mark-recapture studies of closed populations. Biometrics 48:567-576.

Hirst, D. 1994. An improved removal method for estimating animal abundance. Biometrics 50:501-505.

Hoffman, A. 1993. Quantifying selection in wild populations using known-fate and mark-recapture designs. Ph.D. dissertation. University of Washington, Seattle, WA. 259 p.

Iwamoto, R.N., W.D.Muir, B.P. Sandford, K.W. McIntyre, D.A. Frost, J.G. Williams, S.G. Smith, and J.R. Skalski. 1994. Survival estimates for the passage of juvenile chinook salmon through Snake River dams and reservoirs. Annual report to Bonneville Power Administration, Portland, OR. Contract DE-AI79-93BP10891, Project 93-29, 126 p. + appendices. (Available from Northwest Fisheries Science Center, 2725 Montlake Blvd. E., Seattle, WA 98112-2097.)

Jolly, G.M. 1965. Explicit estimates from capture-recapture data with both death and immigration--stochastic model. Biometrika 52:225-247.

Kalbfleisch, J.D. and D.A. Sprott. 1970. Application of likelihood methods to models involving large numbers of parameters. Journal of the Royal Statistical Society, Series B 32:175-208.

Muir, W.D., S.G.Smith, R.N. Iwamoto, D.J. Kamikawa, K.W. McIntyre, E.E. Hockersmith, B.P. Sandford, P.A. Ocker, T.E. Ruehle, J.G. Williams, and J.R. Skalski. 1995. Survival estimates for the passage of juvenile salmonids through Snake River dams and reservoirs. Annual report to Bonneville Power Administration, Portland, OR. Contract DE-AI79-93BP10891, Project 93-29, 162 p. + appendices. (Available from Northwest Fisheries Science Center, 2725 Montlake Blvd. E., Seattle, WA 98112-2097.)

Seber, G.A.F. 1965. A note on the multiple recapture census. Biometrika 52:249-259.

Skalski, J.R., A. Hoffman, and S.G. Smith. 1993. Testing the significance of individual- and cohort-level covariates in animal survival studies. *In* The use of marked individuals in the study of bird populations dynamics: models, methods, and software. *Edited b*y J.D. Lebreton and P.M. North. Birkhauser Verlag, Basel. p. 9-28

Smith, S.G. 1991. Assessing hazards in wild populations using auxiliary variables in tag-release models. Ph.D. dissertation. University of Washington, Seattle, WA. 351 p.

Smith, S.G., J.R. Skalski, J.W. Schlechte, A. Hoffman, and V. Cassen. 1994. Statistical survival analysis of fish and wildlife tagging studies. SURPH.1 manual. (Available from Columbia Basin Research, Box 358218, University of Washington, Seattle, WA 98195-8218.)